

# Ergodicity of Markov chain Monte Carlo with reversible proposal

Kengo KAMATANI\*

Graduate School of Engineering Science and Center for Mathematical Modeling and Data Science, Osaka University

Dated: February 10, 2016

## Abstract

We describe ergodic properties of some Metropolis-Hastings (MH) algorithms for heavy-tailed target distributions. The analysis usually falls into sub-geometric ergodicity framework but we prove that the mixed preconditioned Crank-Nicolson (MpCN) algorithm has geometric ergodicity even for heavy-tailed target distributions. This useful property comes from the fact that the MpCN algorithm becomes a random-walk Metropolis algorithm under suitable transformation.

**Keywords:** Markov chain; Ergodicity; Monte Carlo; Regular variation;

**MSC2010:** 65C05; 65C40; 60J05

## 1 Introduction

In Bayesian analysis, direct calculation of integral is usually quite difficult especially for high-dimension and/or heavy-tailed target distributions. Markov chain Monte Carlo (MCMC) methods such as Metropolis-Hastings (MH) algorithm provides a useful recipe for the approximation of the integral.

Ergodic properties for heavy-tailed case were handled mostly by sub-geometric drift condition (see e.g. Tuominen and Tweedie [1994], Jarner and Roberts [2002], Fort and Moulines [2003], Douc et al. [2004]) since most MCMC do not satisfy geometric drift condition. Application of sub-geometric drift condition to MCMC includes Fort and Moulines [2000], Jarner and Tweedie [2003], Douc et al. [2004], Jarner and Roberts [2007], and Atchadé and Fort [2010]. On the other hand some other MCMC can be geometrically ergodic. This includes independent sampler and position dependent variance MH algorithm on  $\mathbb{R}$  such as Dutta [2010], Livingstone [2015]. Note that independent sampler is very sensitive for the choice of the proposal distribution, and position dependent methods have difficulty in high-dimension which may negatively affect ergodic properties.

In this paper, we consider geometric ergodicity for multidimensional heavy-tailed and light-tailed target distributions. Recently the mixed preconditioned Crank-Nicolson (MpCN) algorithm was considered in Kamatani [2014]. The method has good high-dimensional properties even for heavy-tailed target distributions. As the number of dimension  $d \rightarrow \infty$ , the number of iteration until convergence is  $O(d)$  whereas the random-walk Metropolis one is  $O(d^2)$ . To prove ergodicity, we provide the key property, the **random-walk Metropolis** property for the MpCN kernel: The MpCN kernel becomes a random-walk Metropolis kernel under suitable transformation. Thus MpCN is considered to be an extreme case of variable transformation methods (see Kamatani [2009] and Johnson and Geyer [2012]). By using this fact, it is rather straightforward to show geometric ergodicity for fairly general class of target distributions in  $\mathbb{R}^d$ .

The main result is summarized in the next theorem. The formal definition of the MpCN kernel is in Section 2.1 and the proof is deferred to Sections 3.2 and 3.3.

---

\*Supported in part by Grant-in-Aid for Young Scientists (B) 24740062 and CREST JST.

**Theorem 1.** *The MpCN kernel is geometrically ergodic for the target probability distribution  $\Pi(dx) = \pi(x)dx$  on  $\mathbb{R}^d$  such that*

**Heavy-tailed class**  $\pi(x)$  *is strictly positive continuous function such that*

$$\lim_{r \rightarrow \infty} \frac{\pi(rx)}{\pi(r1)} = \|x\|^{-\alpha}$$

*for some  $\alpha > d$  where the above convergence is locally uniform in  $x$ .*

**Light-tailed class**  $\pi(x)$  *is strictly positive differentiable function such that*

$$\lim_{r \rightarrow \infty} \frac{\pi(rsx)}{\pi(rx)} = \begin{cases} 0 & \text{if } 1 < s \\ +\infty & \text{if } 1 > s \end{cases}$$

*for any  $x \neq 0$ , and satisfies a curvature condition*

$$\limsup_{x \rightarrow \infty} \left\langle \frac{x}{\|x\|}, \frac{\nabla \log \pi(x)}{\|\nabla \log \pi(x)\|} \right\rangle < 0$$

The heavy-tail class includes (a) polynomial target densities considered in Jarner and Roberts [2007] (Section 3.3), and the light-tailed class includes (b) super-exponential densities in Jarner and Hansen [2000] (Section 4) and (c) exponential densities in Fort and Moulines [2000] (Assumption D). Note that the random-walk Metropolis algorithm is geometrically ergodic only for (b) (Theorem 4.3 of Jarner and Hansen [2000]).

The rest of the paper is organized as follows. In Section 2 MpCN algorithm is introduced as a MH kernel with reversible proposal. In this section, the random-walk Metropolis property is defined and proved that the MpCN kernel has the property. Section 3 provides ergodic properties of MpCN kernel.

We finish the section with notation that will be used through the paper.  $N_d(\mu, \Sigma)$  is the  $d$ -dimensional normal distribution with mean  $\mu$  and variance covariance matrix  $\Sigma$ , and  $\phi_d(x)$  is the density of  $N_d(0, I_d)$  where  $I_d$  is the  $d \times d$  identity matrix.  $\mathcal{L}(X)$  is the law of the random variable  $X$ .

## 2 The MpCN kernel

In this section we describe the mixed preconditioned Crank-Nicolson (MpCN) algorithm as an MH kernel with reversible proposal kernel. For general background on Markov chain we refer to Nummelin [1984] and Meyn and Tweedie [2009] and MCMC to Tierney [1994], and Brooks et al. [2011].

### 2.1 Metropolis-Hastings kernels with reversible proposals

Let  $(E, \mathcal{E})$  be a measurable space and let  $P$  be a (probability) transition kernel and  $\Pi(dx)$  be a probability measure. The transition kernel  $P$  is called  $\nu$ -**reversible** if  $\nu(dx)P(x, dy) = \nu(dy)P(y, dx)$  for a  $\sigma$ -finite measure  $\nu$ . Let  $\tilde{\Pi}$  be a  $\sigma$ -finite measure on  $(E, \mathcal{E})$ , and  $\pi$  and  $\tilde{\pi}$  be the densities of  $\Pi$  and  $\tilde{\Pi}$  with respect to a  $\sigma$ -finite measure. If transition kernel  $Q$  is  $\tilde{\Pi}$ -reversible, **Metropolis-Hastings (MH) kernel**  $P$  (with reversible proposal) is defined by

$$P(x, dy) = Q(x, dy)\alpha(x, y) + \delta_x(dy) \left( 1 - \int_{z \in E} Q(x, dz)\alpha(x, z) \right)$$

where

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)\tilde{\pi}(x)}{\pi(x)\tilde{\pi}(y)} \right\}.$$

We call  $Q$  the proposal kernel of  $P$ . MH kernel is  $\Pi$ -reversible.

In this paper, three MH kernels on Euclidean space will be studied. Assume  $d \geq 2$ . Let  $S^{d-1}$  be the unit sphere in  $\mathbb{R}^d$  defined by  $\|x\| = 1$  where  $\langle \cdot, \cdot \rangle$  is the Euclidean inner product and  $\|x\| = \langle x, x \rangle^{1/2}$ . A probability measure  $\Gamma$  on  $\mathbb{R}^d$  is called symmetric about the origin if  $\Gamma(A) = \Gamma(-A)$  for any Borel set  $A$  where  $-A = \{-x; x \in A\}$ .

**Definition 2.1** (RWM kernel). *The random-walk Metropolis kernel uses  $Q(x, dx^*) = \Gamma(dx^* - x)$  where the probability distribution  $\Gamma$  is symmetric about the origin. In this case  $\tilde{\Pi}$  is the Lebesgue measure. Its ergodic properties were studied in Mengersen and Tweedie [1996], Roberts and Tweedie [1996], Jarner and Hansen [2000] and Fort and Moulines [2000].*

**Definition 2.2** (pCN kernel). *The preconditioned Crank Nicolson (pCN) kernel (Beskos et al. [2008]) uses*

$$x^* \leftarrow \rho^{1/2}x + (1 - \rho)^{1/2}w, \quad w \sim N_d(0, I_d)$$

*as the proposal kernel. In this case  $\tilde{\Pi}$  is the standard normal distribution. This method has mainly been studied by high-dimensional analysis, see e.g., Hairer et al. [2014].*

To obtain a better mixing property, we consider scale mixture version of the pCN kernel. See Kamatani [2014] for more background and high-dimensional asymptotic theory.

**Definition 2.3** (MpCN kernel). *The mixed pCN (MpCN) kernel (Kamatani [2014]) uses*

$$r \sim \text{Gamma}(d/2, \|x\|^2/2), \quad w \sim N_d(0, I_d), \text{ and} \\ x^* \leftarrow \rho^{1/2}x + (1 - \rho)^{1/2}r^{-1/2}w.$$

*as the proposal kernel. In this case  $\tilde{\Pi}(dx) = \|x\|^{-d}dx$ .*

In the above,  $\text{Gamma}(\nu, \alpha)$  is the Gamma distribution with the shape parameter  $\nu$  and the scale parameter  $\alpha$  with the probability distribution function  $\propto x^{\nu-1} \exp(-\alpha x)$ . We usually set  $\rho = 0.8$ . Obviously, the proposal kernels for RWM and pCN are reversible and it is also true for MpCN kernel. See Lemma 2.1 of Kamatani [2014] for the proof.

Since  $\|\tilde{w}\|^2$  follows the chi-squared distribution  $\text{Gamma}(d/2, 1/2)$ , we have another useful expression

$$x^* \leftarrow \rho^{1/2}x + (1 - \rho)^{1/2}\|x\| \frac{w}{\|\tilde{w}\|} \quad (2.1)$$

for the proposal of the MpCN kernel, where  $w, \tilde{w} \sim N_d(0, I_d)$  are independent. By this notation,

$$\|x^*\|^2 = \|x\|^2 \left( \rho + 2\sqrt{\rho(1-\rho)} \frac{\|w\|}{\|\tilde{w}\|} v + (1 - \rho) \frac{\|w\|^2}{\|\tilde{w}\|^2} \right) \quad (2.2)$$

where  $v = \langle w/\|w\|, x/\|x\| \rangle$ . The law of  $v$  is the first element of the uniform distribution on  $S^{d-1}$  and it is independent from  $\|w\|$  and  $\|\tilde{w}\|$ . Therefore the law of  $\|x^*\|/\|x\|$  does not depend on  $x$ . In particular, the law of

$$\xi(x) = \log(\|x^*\|^2) - \log(\|x\|^2) = \log \left( \frac{\|x^*\|^2}{\|x\|^2} \right) \quad (2.3)$$

does not depend on  $x$ .

## 2.2 Random walk property

In this section we will present the random-walk Metropolis (RWM) property of the MpCN kernel which is the key for the proof of ergodicity in Section 3. Let  $\Psi^{-1}A = \{x; \Psi(x) \in A\}$ .

**Definition 2.4.** A transition kernel  $P$  on  $(E, \mathcal{E})$  has the **random-walk property** with respect to  $\Psi : E \rightarrow \mathbb{R}^d$  if there exists a probability distribution  $\Gamma$  which is symmetric about the origin such that

$$P(y, \Psi^{-1}A) = \Gamma(A - x)$$

for all  $x \in \mathbb{R}^d, y \in E$  and Borel set  $A$  such that  $\Psi(y) = x$ . The MH kernel  $P$  has the **RWM property** with respect to  $\Psi$  if its proposal kernel  $Q$  has the random-walk property with respect to  $\Psi$ .

A few methods with this property have been proposed in the literature, including multiplicative random walk in Dellaportas and Roberts [2003] and transformation method in Johnson and Geyer [2012].

**Proposition 2.1.** The law of  $\xi(x)$  in (2.3) is symmetric about the origin and does not depend on  $x$ . In particular, the MpCN kernel has the RWM property with respect to  $\Psi(x) = \log(\|x\|^2)$ .

*Proof.* By expression (2.2), the law of  $\xi(x)$  and  $\xi(\tilde{w})$  are the same as described above. Note that

$$\xi(\tilde{w}) = \log \left( \|\rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w\|^2 \right) - \log(\|\tilde{w}\|^2).$$

Moreover, there exists exchangeability  $\mathcal{L}(\tilde{w}, \rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w) = \mathcal{L}(\rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w, \tilde{w})$ . Therefore the law of  $\xi(\tilde{w})$  is symmetric about the origin since

$$\begin{aligned} \mathbb{P}_x(\xi(\tilde{w}) > t) &= \mathbb{P}_x(\log \|\rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w\|^2 - \log \|\tilde{w}\|^2 > t) \\ &= \mathbb{P}_x(\log \|\tilde{w}\|^2 - \log \|\rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w\|^2 > t) = \mathbb{P}_x(\xi(\tilde{w}) < -t) \quad (t \in \mathbb{R}). \end{aligned}$$

Thus the claim holds by putting  $\Gamma = \mathcal{L}(\xi(\tilde{w})) = \mathcal{L}(\log \|\rho^{1/2}\tilde{w} + (1-\rho)^{1/2}w\|^2 - \log \|\tilde{w}\|^2)$ .  $\square$

### 3 Ergodicity

We have introduced the MpCN kernel in Section 2 as an extension to the pCN kernel and showed it to have the RWM property. For this reason, the ergodic properties of the MpCN kernel can be derived in the same way as that of the RWM kernel. We consider heavy-tailed target distributions in Section 3.2 and light-tailed target distributions in Section 3.3. We prepare Section 3.1 for necessary condition for geometric ergodicity. We will conclude that unlike the RWM and pCN kernels, the MpCN kernel is geometrically ergodic for very wide class of target distributions.

I will begin by reviewing a few elementary properties of transition kernels. Our notation and terminologies generally follow those of Meyn and Tweedie [2009]. Let  $P(x, dy)$  be a transition kernel on a measurable space  $(E, \mathcal{E})$ . We define

$$Ph(x) = \int_y P(x, dy)h(y), \quad (\nu P)(dy) = \int_x \nu(dx)P(x, dy)$$

for any measurable function  $h(x)$  and signed measure  $\nu$  if the right-hand side exists. A probability measure  $\Pi$  is called the **invariant probability measure** if  $\Pi P = \Pi$  and  $P$  is called  $\Pi$ -invariant. Let  $P^0(x, dy) = I(x, dy) := \delta_x(dy)$  and  $P^{k+1}(x, dy) = \int_z P(x, dz)P^k(z, dy)$  ( $k \geq 0$ ). The kernel  $P$  is called  **$\Pi$ -irreducible** if  $\Pi$  is absolutely continuous with respect to  $\sum_{k=1}^{\infty} P^k(x, \cdot)$  for any  $x \in E$ . A set  $C \in \mathcal{E}$  is called **small set** if

$$P^k(x, \cdot) \geq \delta \nu \quad (x \in C) \tag{3.1}$$

for some  $k \in \mathbb{N}$ ,  $\delta \in (0, 1)$ , and a probability measure  $\nu$ . We require usual assumptions throughout in this paper: (a)  $P$  is  $\Pi$ -irreducible (b)  $P$  is  $\Pi$ -invariant (c) there exists a small set  $C \in \mathcal{E}$  such that  $\Pi(C) > 0$  and (d)  $\Pi$  is not singular, that is,  $\Pi(\{x\}) < 1$  for  $x \in E$ . Note that if  $\mathcal{E}$  is countably generated, (c) comes from (a) (Proposition 2.6 of Nummelin [1984]).

Let  $V : E \rightarrow [1, \infty]$  be a function such that  $V(x) < \infty$  for  $\Pi$ -a.s. The transition kernel  $P$  is said to have the **geometric drift condition** if there is a small set  $C$ ,  $\gamma \in (0, 1)$  and  $b < \infty$

$$PV \leq \gamma V + b1_C. \quad (3.2)$$

The condition is extensively studied in the past few decades. In particular, if the above condition is satisfied, and also there exists a small set that satisfies (3.1) for  $k = 1$ , then  $P$  is **geometrically ergodic**, that is

$$\|P^n - \Pi\|_V \leq c\gamma^n V(x)$$

where  $\|\nu\|_V = \sup_{f: |f| \leq V} |\int f(x)\nu(dx)|$  for a signed measure  $\nu$  and  $c$  is a constant (See Theorem 15.0.1 of Meyn and Tweedie [2009]). Moreover, geometric ergodicity implies geometric drift condition if the conditions (a) and (c) are satisfied (See Theorem 16.0.1 of Meyn and Tweedie [2009]).

### 3.1 Necessary condition for ergodicity

In this section we introduce necessary condition for geometric ergodicity for random-walk type kernels (RWM and MpCN) and MH kernel with ergodic proposal kernel (pCN). Let  $(E, d)$  be a pseudometric space, that is,  $d(x, y) \geq 0$ ,  $d(x, y) = d(y, x)$  and  $d(x, z) \leq d(x, y) + d(y, z)$ . Let  $B_r(x) = \{y; d(x, y) < r\}$ . Let  $\mathcal{E}$  be its Borel  $\sigma$ -algebra generated by the pseudometric topology. Fix  $x^* \in E$ . The RWM kernel, and the MpCN kernel after transform  $\Psi$  satisfy the following property.

**Assumption 1.** *For any  $\epsilon > 0$ , there exists  $r > 0$  such that  $P(x, B_r(x)) > 1 - \epsilon$  for any  $x \in E$ .*

The following proposition, due to Jarner and Tweedie [2003], gives necessary condition for ergodicity. This says that if  $P$  is geometrically ergodic, the target distribution has exponential tail. We give a proof for the sake of convenience of the reader.

**Proposition 3.1** (Jarner and Tweedie [2003]). *Assume Assumption 1. If  $P$  satisfies geometric ergodicity then there exists  $\delta > 0$  such that*

$$\int_{x \in E} \exp(\delta d(x^*, x)) \Pi(dx) < \infty.$$

To prove the proposition, we need two simple lemmas. The first lemma says that small set is “small”. The second lemma says that under geometric ergodicity, there is a small set “large enough”.

**Lemma 3.1.** *Under Assumption 1, any small set is bounded.*

*Proof.* Assume by contradiction that there is an unbounded small set  $C$  such that (3.1). Choose  $r > 0$  so that  $P(x, B_r(x)) > (1 - \delta/2)^{1/k}$ . Then by definition, by putting  $x_0 = x$ ,

$$P^k(x, B_{kr}(x)) \geq \int \prod_{l=1}^k I_{B_r(x)}(x_l) P(x_{l-1}, dx_l) > \overbrace{(1 - \delta/2)^{1/k} \cdots (1 - \delta/2)^{1/k}}^k = 1 - \delta/2$$

and thus  $P^k(x, B_{kr}(x)^c) \leq \delta/2$ . Since  $C$  is unbounded, we can choose  $x_i \in C$  ( $i = 1, 2$ ) so that  $B_{kr}(x_1) \cap B_{kr}(x_2) = \emptyset$ . Then

$$P^k(x_i, B_{kr}(x_i)^c) \geq \delta \nu(B_{kr}(x_i)^c)$$

and hence  $\nu(B_{kr}(x_i)) = 1 - \nu(B_{kr}(x_i)^c) \geq 1 - \delta^{-1} P^k(x_i, B_{kr}(x_i)^c) > 1/2$  for  $i = 1, 2$ . This would imply  $\nu(E) \geq \nu(B_{kr}(x_1)) + \nu(B_{kr}(x_2)) > 1$  which is a contradiction. Thus any small set is bounded.  $\square$

**Lemma 3.2.** *If transition kernel  $P$  satisfies (3.2), then there exists  $s > 1$  such that*

$$\sum_{n=0}^{\infty} s^n (I_{C^c} P)^n(x, E)$$

*is  $\Pi$ -integrable.*

*Proof.* By (3.2),

$$(I_{C^c} P)V = I_{C^c}(PV) \leq I_{C^c}(\gamma V + b1_C) = I_{C^c}\gamma V \leq \gamma V.$$

Then by choosing  $s > 1$  so that  $s\gamma < 1$ , we have

$$\sum_{n=0}^{\infty} s^n (I_{C^c} P)^n(x, E) \leq \sum_{n=0}^{\infty} s^n (I_{C^c} P)^n V \leq \sum_{n=0}^{\infty} s^n \gamma^n V \leq \frac{1}{1 - s\gamma} V.$$

By assumption,  $P$  is irreducible and  $\Pi$ -invariant. By Theorem 14.3.7 of Meyn and Tweedie [2009], the drift function  $V$  in (3.2) is  $\Pi$ -integrable. Hence the left-hand side is also  $\Pi$ -integrable  $\square$

*Proof of Proposition 3.1.* By Lemma 3.1, small set  $C$  is bounded. We choose  $r > 0$  such a way that  $C \subset B_r(x^*)$  and  $P(x, B_r(x)) > s^{-1}$  ( $x \in E$ ) where  $s > 1$  is as in Lemma 3.2. If  $x_0 \notin B_{nr}(x^*)$  and if  $x_n \in B_r(x_{n-1})$  ( $n \geq 1$ ), then  $x_0, \dots, x_{n-1} \notin B_r(x^*)$ . Therefore

$$(I_{C^c} P)^n(x_0, E) \geq (I_{B_r(x^*)^c} P)^n(x_0, E) \geq \int \prod_{m=1}^n I_{B_r(x_{m-1})}(x_m) P(x_{m-1}, dx_m) \geq (\inf_x P(x, B_r(x)))^n =: \eta^n$$

where  $\eta > s^{-1}$ . Thus

$$s^{n(x)} (I_{C^c} P)^{n(x)}(x, E) \geq (s\eta)^{n(x)} = \exp(n(x) \log(s\eta))$$

where  $n(x) = \lceil d(x^*, x)/r \rceil \geq d(x^*, x)/r - 1$  where  $\lceil t \rceil$  is the integer part of  $t > 0$ . Therefore we can find  $c > 0$  and  $\delta > 0$  such that

$$s^{n(x)} (I_{C^c} P)^{n(x)}(x, E) \geq c \exp(\delta d(x^*, x)).$$

Since the left-hand side is  $\Pi$ -integrable by Lemma 3.2, the right-hand side is also  $\Pi$ -integrable.  $\square$

By Proposition 3.1, RWM kernel is geometrically ergodic only if  $\Pi$  has a light-tailed density. The MpCN kernel has the same property but after the projection  $x \mapsto \log \|x\|^2$ . The requirement of the MpCN kernel is that  $\Pi$  has a polynomial-tailed density which is much weaker condition compared to the RWM kernel.

**Corollary 3.1** (Jarner and Tweedie [2003]). *The random-walk Metropolis kernel on  $\mathbb{R}^d$  satisfy Assumption 1 for Euclidean metric  $d(x, y)$ . Thus if the kernel is geometrically ergodic, by taking  $x^* = 0$ ,*

$$\int \exp(\delta \|x\|) \Pi(dx) < \infty \tag{3.3}$$

*for some  $\delta > 0$ , where  $\|\cdot\|$  is Euclidean norm.*

**Corollary 3.2.** *The MpCN kernel on  $\mathbb{R}^d$  satisfy Assumption 1 for  $d(x, y) = |\log(\|x\|^2) - \log(\|y\|^2)|$  on where  $\|\cdot\|$  is Euclidean norm. Thus if the kernel is geometrically ergodic, by taking  $x^* \in S^{d-1}$ ,*

$$\int \exp(\delta |\log(\|x\|^2)|) \Pi(dx) < \infty$$

*for some  $\delta > 0$ . In particular,*

$$\int \|x\|^\delta \Pi(dx) < \infty.$$

Proposition 3.1 is useful for MH kernels with transient proposal kernels, but may not be useful for those with ergodic proposal kernels. In order to study necessary condition for the latter case, we need to estimate the acceptance probability. There is a useful result due to Mengersen and Tweedie [1996], Roberts and Tweedie [1996].

**Proposition 3.2** (Roberts and Tweedie [1996]). *If  $P$  is geometrically ergodic, then  $\Pi$ -ess sup  $P(x, \{x\}) < 1$ .*

*Proof.* Let  $E' = \{x; V(x) < \infty\}$ . To obtain a contradiction, suppose  $\Pi$ -ess sup  $P(x, \{x\}) = 1$ . Choose  $x_n \in E'$  so that  $P(x_n, \{x_n\}) \geq 1 - n^{-1}$ . Any small set  $C$  only includes finitely many elements of  $\{x_n\}_n$ . Otherwise, if (3.1) is satisfied, then

$$\delta\nu(\{x_n; n \geq N\}^c) \leq \delta\nu(\{x_N\}^c) \leq P^k(x_N, \{x_N\}^c) \leq 1 - \left(1 - \frac{1}{N}\right)^k$$

for each  $x_N \in C$ . Taking  $N \rightarrow \infty$  we have  $\delta = 0$  and hence this contradicts  $C$  is a small set.

By geometric ergodicity, (3.2) is satisfied. Choose  $x_n$  as above such that  $x_n \notin C$ . Then

$$\left(1 - \frac{1}{n}\right) V(x_n) \leq P(x_n, \{x_n\}) V(x_n) \leq PV(x_n) \leq \gamma V(x_n).$$

By taking  $n \rightarrow \infty$ ,  $\gamma = 1$  and hence this contradicts our assumption for geometric ergodicity of  $P$ .  $\square$

We state a necessary condition for ergodicity for the pCN kernel as a corollary of Proposition 3.2. It says that the pCN kernel requires even lighter-tailed density for the target distribution than the RWM kernel.

**Corollary 3.3.** *Suppose that  $\Pi$  has a probability density  $\pi(x)$ . For each  $r > 0$ , let*

$$C_r = r^{-2} \sup_{\rho r \leq \|x\|, \|y\| \leq r} |\log \pi(x) - \log \pi(y)|.$$

*If the pCN kernel is geometrically ergodic, then  $\liminf_{r \rightarrow \infty} C_r \geq (1 - \rho)/2$ .*

*Proof.* Write  $x^* = \rho^{1/2}x + (1 - \rho)^{1/2}w$ . Assume that the pCN kernel is geometrically ergodic. By Proposition 3.2,

$$\delta < 1 - P(x, \{x\}) = \int \min \left\{ 1, \frac{\pi(x^*)\phi_d(x)}{\pi(x)\phi_d(x^*)} \right\} \phi_d(w) dw \quad (3.4)$$

for (Leb) a.s.  $x$  for some  $\delta > 0$ . By triangular inequality, for sufficiently large  $r = \|x\|$ , we have  $\rho r \leq \|x^*\| \leq r$  since

$$\|x^*\| = \|\rho^{1/2}x + (1 - \rho)^{1/2}w\| = \rho^{1/2}\|x\| + o(\|x\|)$$

with an obvious inequality  $\rho < \rho^{1/2} < 1$ . Thus for each  $w \in \mathbb{R}^d$ ,

$$\begin{aligned} \log \left( \frac{\pi(x^*)\phi_d(x)}{\pi(x)\phi_d(x^*)} \right) &= \{\log \pi(x^*) - \log \pi(x)\} + \{\log \phi_d(x) - \log \phi_d(x^*)\} \\ &= \{\log \pi(x^*) - \log \pi(x)\} + \left\{ -\frac{\|x\|^2}{2} + \frac{\rho\|x\|^2 + 2\sqrt{\rho(1-\rho)}\langle x, w \rangle + (1-\rho)\|w\|^2}{2} \right\} \\ &\leq C_r r^2 - \frac{1-\rho}{2} r^2 + O(r) \\ &= r^2 \left\{ C_r - \frac{1-\rho}{2} + O(r^{-1}) \right\}. \end{aligned}$$

Therefore if  $\liminf C_r < (1 - \rho)/2$ , we can choose a sequence of  $r = r_n = \|x_n\|$  such that the right-hand side of the above tends to 0. By Lebesgue's dominated convergence theorem, the right-hand side of (3.4) converges to 0 for this sequence, which is a contradiction. Thus  $\liminf C_r \geq (1 - \rho)/2$ .  $\square$

### 3.2 Ergodicity for regular varying function

We prove geometric ergodicity in terms of regularly varying property. For introductory literature to regularly varying functions we refer the reader to the books Bingham et al. [1989], Resnick [2008]. The theory of regularly varying functions provides a framework for heavy-tail analysis. For one dimensional case, a positive function  $h(r)$  on  $(0, \infty)$  is called regularly varying if  $\lim_{r \rightarrow \infty} h(rx)/h(r) = \lambda(x)$  for some positive finite valued function  $\lambda$ . We consider multidimensional version. We denote  $a(r, x) \xrightarrow{ucp} a(x)$  ( $r \rightarrow \infty$ ) if for any  $x \in \mathbb{R}^d \setminus \{0\}$  there exists a compact set  $K \ni x$  such that  $\lim_{r \rightarrow \infty} \sup_{y \in K} |a(r, y) - a(y)| = 0$ .

**Definition 3.1.** *Positive valued function  $h(x)$  ( $x \in \mathbb{R}^d$ ) is called symmetrically regularly varying if*

$$\frac{h(rx)}{h(r1)} \xrightarrow{ucp} \lambda(x) \quad (r \rightarrow \infty)$$

for some  $\lambda : \mathbb{R} \rightarrow (0, \infty)$  such that  $\lambda(x) = 1$  for any  $x \in S^{d-1}$  where  $1 = (1, \dots, 1) \in \mathbb{R}^d$ .

This class includes many functions such as polynomial target densities considered in Jarner and Roberts [2007] (Section 3.3). This class inherits useful properties from one dimensional regularly varying function:  $\lambda(x) = \|x\|^{-\alpha}$  for the **exponent of variation**  $-\alpha \in \mathbb{R}$  (p277 of Resnick [2008]). Note that symmetry of  $\lambda(x)$  is crucial in our proof. It is not obvious to construct a simple sufficient condition for geometric ergodicity for non-symmetric case.

Assume that  $\Pi$  has the density  $\pi(x)$ . Before stating the main result of this section we prove simple lemma for integrability of the regularly varying function.

**Lemma 3.3.** *If  $\int \|x\|^\delta \Pi(dx) < \infty$  for some  $\delta > 0$ , then the exponent of variation  $-\alpha$  of the symmetrically regularly varying function  $\pi(x)$  satisfies  $\alpha > d$ .*

*Proof.* For  $x > 0$ , let  $h(r) = \int_{\xi \in S^{d-1}} \pi(r\xi) r^{d-1} d\xi$ . Then  $h(r)$  is a regularly varying function with exponent of variation  $-\alpha + d - 1$  by local uniform convergence property since

$$\frac{h(rs)}{h(r)} = \frac{h(rs)/\pi(r1)}{h(r)/\pi(r1)} = \frac{\int_{\xi \in S^{d-1}} \pi(rs\xi)/\pi(r1) (rs)^{d-1} d\xi}{\int_{\xi \in S^{d-1}} \pi(r\xi)/\pi(r1) r^{d-1} d\xi} \rightarrow \frac{\int_{\xi \in S^{d-1}} \lambda(s\xi) (rs)^{d-1} d\xi}{\int_{\xi \in S^{d-1}} \lambda(\xi) r^{d-1} d\xi} = s^{-\alpha+d-1} \quad (r \rightarrow \infty).$$

Therefore, by Potter bounds (Theorem 1.5.6 (iii) of Bingham et al. [1989]), if  $\beta - \alpha + d - 1 > -1$ , then  $\int_1^\infty r^\beta h(r) dr = \infty$ . Since  $\int_1^\infty r^\delta h(r) dr = \int_{\|x\| > 1} \|x\|^\delta \Pi(dx) < \infty$ , we have  $\delta - \alpha + d - 1 \leq -1$ . Thus  $\delta + d \leq \alpha$ , and hence  $d < \alpha$ .  $\square$

**Proposition 3.3.** *If  $\pi(x)$  is symmetrically regularly varying function, then RWM kernel and pCN kernel do not have geometric ergodicity.*

*Proof.* Let  $h(r)$  be as in the previous lemma. Then  $h(r)$  is a regularly varying function and hence  $\int e^{s\|x\|} \Pi(dx) = \int e^{sr} h(r) dr = +\infty$  (Theorem 1.5.6 (iii) of Bingham et al. [1989]). Hence RWM kernel does not have geometric ergodicity by Corollary 3.1.

Next we consider pCN kernel. By local uniform convergence property,

$$\begin{aligned} r^2 C_r &= \sup_{\rho \leq \|x\|, \|y\| \leq 1} |\log \pi(rx) - \log \pi(ry)| \\ &= \sup_{\rho \leq \|x\|, \|y\| \leq 1} \left| \left( \log \frac{\pi(rx)}{\pi(r1)} - \log \lambda(x) \right) - \left( \log \frac{\pi(ry)}{\pi(r1)} - \log \lambda(y) \right) + (\log \lambda(x) - \log \lambda(y)) \right| \\ &\leq 2 \sup_{\rho \leq \|x\| \leq 1} \left| \log \frac{\pi(rx)}{\pi(r1)} - \log \lambda(x) \right| + \sup_{\rho \leq \|x\|, \|y\| \leq 1} |\log \lambda(x) - \log \lambda(y)| \\ &\rightarrow \sup_{\rho \leq \|x\|, \|y\| \leq 1} |\log \lambda(x) - \log \lambda(y)| < \infty \quad (t \rightarrow \infty). \end{aligned}$$

Thus  $C_r = o(1)$ , and hence pCN kernel does not have geometric ergodicity by Corollary 3.3.  $\square$



By Corollary 3.2, the MpCN kernel is geometrically ergodic only if  $\Pi$  has a polynomial tail. The following proposition states the converse.

**Proposition 3.4.** *Assume  $\pi(x)$  is strictly positive continuous symmetrically regularly varying function. Then the MpCN kernel is geometrically ergodic if and only if  $\int \|x\|^\delta \Pi(dx) < \infty$  for some  $\delta > 0$ .*

*Proof.* We use expression in (2.1) and (2.3). Let  $q(x) = \pi(x)\|x\|^d$  and let  $V(x) = \begin{cases} q(x)^{-s} & x \neq 0 \\ +\infty & x = 0 \end{cases}$  for  $s \in (0, 1)$ . Then  $V(x)$  is bounded on  $C = \{x; r \leq \|x\| \leq r^{-1}\}$  for any  $r \in (0, 1)$ , and  $C$  is a small set for the MpCN kernel. To prove (3.2), it is sufficient to show

$$\limsup_{x \rightarrow 0} \frac{PV(x) - V(x)}{V(x)} < 0, \quad \limsup_{x \rightarrow \infty} \frac{PV(x) - V(x)}{V(x)} < 0. \quad (3.5)$$

Observe that

$$\frac{PV(x) - V(x)}{V(x)} = \mathbb{E}_x \left[ \left\{ \left( \frac{q(x^*)}{q(x)} \right)^{-s} - 1 \right\} \min \left\{ 1, \frac{q(x^*)}{q(x)} \right\} \right] \quad (3.6)$$

and the integrand is uniformly bounded. Since  $\pi(x)$  is continuous at 0, for each  $w, \tilde{w}$ ,

$$1 = \lim_{x \rightarrow 0} \frac{\pi(x^*)}{\pi(x)} = \lim_{x \rightarrow 0} \frac{q(x^*)}{q(x)} \exp \left( -\frac{d}{2} \xi(x) \right).$$

Since the law of  $\xi(x)$  is independent of  $x$ , we simply write  $\xi$  for  $\xi(x)$ . Then by Slutsky's theorem,  $q(x^*)/q(x)$  converges in law to  $\exp(d\xi/2)$  as  $x \rightarrow 0$ . Therefore we have

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{PV(x) - V(x)}{V(x)} &= \mathbb{E} \left[ \left\{ e^{-\frac{ds}{2}\xi} - 1 \right\} \min \left\{ 1, e^{\frac{d}{2}\xi} \right\} \right] \\ &= \mathbb{E} \left[ \left\{ e^{-\frac{ds}{2}\xi} - 1 \right\}, \xi > 0 \right] + \mathbb{E} \left[ e^{\frac{d(1-s)}{2}\xi} \left\{ 1 - e^{\frac{ds}{2}\xi} \right\}, \xi < 0 \right]. \end{aligned}$$

By Proposition 2.1, the law of  $\xi$  is symmetric about the origin. Therefore the above expectation equals to

$$\mathbb{E} \left[ \left\{ e^{-\frac{ds}{2}\xi} - 1 \right\}, \xi > 0 \right] + \mathbb{E} \left[ e^{-\frac{d(1-s)}{2}\xi} \left\{ 1 - e^{-\frac{ds}{2}\xi} \right\}, \xi > 0 \right] = \mathbb{E} \left[ \left( 1 - e^{-\frac{d(1-s)}{2}\xi} \right) \left\{ e^{-\frac{ds}{2}\xi} - 1 \right\}, \xi > 0 \right] < 0$$

since  $\mathbb{P}(\xi > 0) > 0$  and the integrand is negative for any  $\xi > 0$ . Thus the first part of (3.5) is completed.

Now we consider the second part of (3.5). Let  $\eta(x) = (d - \alpha)\xi(x)/2$ . Then for each  $w, \tilde{w}$ ,

$$\frac{q(x^*)}{q(x)} \exp(-\eta(x)) = \frac{\pi(x^*)}{\pi(x)} \exp \left( \frac{\alpha}{2} \xi(x) \right) = \frac{\pi \left( \|x\| \frac{x^*}{\|x\|} \right) / \pi(\|x\| \cdot 1)}{\pi \left( \|x\| \frac{x}{\|x\|} \right) / \pi(\|x\| \cdot 1)} \exp \left( \frac{\alpha}{2} \xi(x) \right).$$

By local uniform convergence of the regular varying function, the right-hand side of the above converges to 1. Since the law of  $\eta(x)$  does not depend on  $x$ , we simply denote it by  $\eta$ . Thus as in the first part of (3.5),  $q(x^*)/q(x)$  converges in law to  $\exp(\eta)$  by Slutsky's theorem as  $x \rightarrow \infty$ . Hence

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{PV(x) - V(x)}{V(x)} &= \mathbb{E} \left[ \left\{ e^{-s\eta} - 1 \right\} \min \{1, e^\eta\} \right] \\ &= \mathbb{E} \left[ \left\{ e^{-s\eta} - 1 \right\}, \eta > 0 \right] + \mathbb{E} \left[ e^{(1-s)\eta} \left\{ 1 - e^{s\eta} \right\}, \eta < 0 \right]. \end{aligned}$$

Since the distribution of  $\eta$  is symmetric, the above expectation equals to

$$\mathbb{E} \left[ \left\{ e^{-s\eta} - 1 \right\}, \eta > 0 \right] + \mathbb{E} \left[ e^{-(1-s)\eta} \left\{ 1 - e^{-s\eta} \right\}, \eta > 0 \right] = \mathbb{E} \left[ \left( 1 - e^{-(1-s)\eta} \right) \left\{ e^{-s\eta} - 1 \right\}, \eta > 0 \right].$$

Since the integrand is negative if  $\eta > 0$ , the claim follows if  $\mathbb{P}(\eta > 0) > 0$ . Since  $\mathbb{P}(\eta \neq 0) = 2\mathbb{P}(\eta > 0)$ , we have geometric ergodicity if  $\mathbb{P}(\eta = 0) < 1$ . However  $\mathbb{P}(\eta = 0) = 1$  is satisfied if and only if  $\alpha = d$ , which contradicts the assumption by Lemma 3.3. Thus the claim follows.  $\square$

### 3.3 Ergodicity for rapidly varying function

In this section we illustrate ergodic property for the MpCN kernel for light-tailed target distributions. We show that the MpCN kernel is geometric ergodicity for any light-tailed target distribution as long as the curvature condition (3.9) is satisfied. On the other hand, as in Corollaries 3.1 and 3.3, super-exponential tail is necessary for the RWM kernel and the pCN kernel. To state the main result, we need a definition for light-tailed distributions.

**Definition 3.2.** A positive function  $h(x)$  ( $x \in \mathbb{R}^d$ ) is rapidly varying if for any  $\xi \in S^{d-1}$ ,  $s > 0$ ,

$$\lim_{r \rightarrow \infty} \frac{h(rs\xi)}{h(r\xi)} = \begin{cases} 0 & \text{if } 1 < s \\ +\infty & \text{if } 1 > s \end{cases}$$

Many light-tailed functions are rapidly varying. For example, super-exponential densities in Jarner and Hansen [2000] (Section 4) and exponential densities in Fort and Moulines [2000] (Assumption D) are rapidly varying functions. See also Johnson and Geyer [2012] for other examples.

**Proposition 3.5.** If  $\pi(x)$  is a continuous strictly positive rapidly varying function, MpCN kernel is geometrically ergodic if and only if  $\Pi\text{-ess sup } P(x, \{x\}) < 1$ .

*Proof.* The only if part follows from Proposition 3.2. The proof will be finished once we show (3.7) in Proposition 3.4 if  $\Pi\text{-ess sup } P(x, \{x\}) < 1$ . We only show the latter inequality in (3.7) since the proof for the former inequality is exactly the same as that of Proposition 3.4. Thanks to  $\lim_{x \rightarrow \pm\infty} e^{-sx} \min\{1, e^x\} = 0$  if

$$\lim_{x \rightarrow \infty} \mathbb{P}_x \left[ \left| \log \frac{q(x^*)}{q(x)} \right| \leq C \right] = 0 \quad (3.7)$$

for any  $C > 0$ , then

$$\lim_{x \rightarrow \infty} \mathbb{E}_x \left[ \left\{ \frac{q(x^*)}{q(x)} \right\}^{-s} \min \left\{ 1, \frac{q(x^*)}{q(x)} \right\} \right] = 0.$$

Therefore by the expression (3.6), the equation (3.7) implies

$$\limsup_{x \rightarrow \infty} \frac{PV(x) - V(x)}{V(x)} = -\liminf_{x \rightarrow \infty} \mathbb{E}_x \left[ \min \left\{ 1, \frac{q(x^*)}{q(x)} \right\} \right] = \limsup_{x \rightarrow \infty} P(x, \{x\}) - 1 = \Pi\text{-ess sup } P(x, \{x\}) - 1$$

where the last equality comes from continuity of  $\pi(x)$ . Thus (3.7) will complete the proof.

By Proposition 2.1, the law of  $\|x^*\|/\|x\| = \exp(\xi/2)$  does not depend on  $x$ . Therefore there exists  $\delta > 0$  for each  $\epsilon > 0$  such that

$$\begin{aligned} \mathbb{P}_x \left[ \left| \log \frac{q(x^*)}{q(x)} \right| \leq C \right] &\leq \mathbb{P}_x \left[ \left| \log \frac{q(x^*)}{q(x)} \right| \leq C, \delta \leq \frac{\|x^*\|}{\|x\|} \leq \delta^{-1} \right] + \mathbb{P}_x \left[ \left\{ \delta \leq \frac{\|x^*\|}{\|x\|} \leq \delta^{-1} \right\}^c \right] \\ &\leq \mathbb{P}_x \left[ \frac{x^*}{\|x\|} \in A(x) \right] + \epsilon \end{aligned}$$

where

$$A(x) = \left\{ y \in \mathbb{R}^d; \left| \log \frac{q(\|x\|y)}{q(x)} \right| \leq C, \delta \leq \|y\| \leq \delta^{-1} \right\}.$$

By expression (2.1),  $x^*/\|x\|$  follows the multivariate Cauchy distribution with shift  $\rho^{1/2}x/\|x\|$  and scale  $1 - \rho$ . Thus the probability distribution function is uniformly bounded, and hence there exists  $c > 0$  such that

$$\mathbb{P}_x \left[ \frac{x^*}{\|x\|} \in A(x) \right] \leq c \text{Leb}(A(x)) = c \int_{\xi \in S^{d-1}} \int_{r \in A(x, \xi)} r^{d-1} dr d\xi \quad (3.8)$$

where

$$A(x, \xi) = \left\{ r > 0; \left| \log \frac{q(\|x\|r\xi)}{q(x)} \right| \leq C, \delta \leq r \leq \delta^{-1} \right\}.$$

By dominated convergence theorem, (3.8) tends to 0 if  $\lim_{x \rightarrow \infty} \text{Leb}(A(x, \xi)) = 0$  for each  $\xi \in S^{d-1}$ . Note that

$$\begin{aligned} A(x, \xi) \times A(x, \xi) &= \left\{ r, s > 0; \left| \log \frac{q(\|x\|r\xi)}{q(x)} \right|, \left| \log \frac{q(\|x\|s\xi)}{q(x)} \right| \leq C, \delta \leq r, s \leq \delta^{-1} \right\} \\ &\subset \left\{ r, s > 0; \left| \log \frac{q(\|x\|r\xi)}{q(\|x\|s\xi)} \right| \leq 2C, \delta \leq r, s \leq \delta^{-1} \right\}. \end{aligned}$$

However by rapidly varying property of  $\pi(x)$ ,

$$\liminf_{x \rightarrow \infty} \left| \log \frac{q(\|x\|r\xi)}{q(\|x\|s\xi)} \right| \geq \liminf_{x \rightarrow \infty} \left| \log \frac{\pi(\|x\|r\xi)}{\pi(\|x\|s\xi)} \right| - \left| \log \frac{(\|x\|r)^d}{(\|x\|s)^d} \right| \rightarrow \infty$$

for each  $\delta \leq r, s \leq \delta^{-1}$  such that  $r \neq s$ , and hence  $\text{Leb}(A(x, \xi)) = \sqrt{\text{Leb}(A(x, \xi) \times A(x, \xi))} \rightarrow 0$ . Thus the claim follows.  $\square$

We state the main result in this section. The curvature condition considered in Jarner and Hansen [2000] is the sufficient condition for geometric ergodicity for the MpCN kernel. The proof follows a similar line of argument to Jarner and Hansen [2000], proof of Theorem 4.3.

**Corollary 3.4.** *If  $\pi(x)$  is differentiable and strictly positive rapidly varying function and if*

$$\limsup_{x \rightarrow \infty} \left\langle \frac{x}{\|x\|}, \frac{\nabla \log \pi(x)}{\|\nabla \log \pi(x)\|} \right\rangle < 0 \quad (3.9)$$

*then MpCN kernel is geometrically ergodic.*

*Proof.* Let  $n(x) = x/\|x\|$  and  $m(x) = \nabla \log \pi(x)/\|\nabla \log \pi(x)\|$ . By assumption, there exists  $\epsilon \in (0, 1)$  and  $M > 0$  such that

$$\langle n(x), m(x) \rangle < -2\epsilon$$

for all  $\|x\| \geq M$ . Let

$$W(x) = \{x - t\|x\|\xi; 0 \leq t \leq \epsilon^2/4, \xi \in S^{d-1}, \|\xi - n(x)\| \leq \epsilon\}.$$

We first prove that

$$(y \in W(x) \text{ and } \|x\| \geq 2M) \Rightarrow \log \frac{\pi(y)}{\pi(x)} > 0. \quad (3.10)$$

If  $y = x - t\|x\|\xi \in W(x)$  then

$$\begin{aligned} \|n(x) - n(y)\|^2 &= 2 - 2\langle n(x), n(y) \rangle = 2 - 2 \left\langle n(x), \frac{n(x) - t\xi}{\|n(x) - t\xi\|} \right\rangle \\ &\leq 2 - 2 \frac{1-t}{1+t} = \frac{4t}{1+t} \leq 4t \leq \epsilon^2 \end{aligned}$$

Since  $1-t \geq 1-\epsilon^2/4 \geq 1/2$ , for  $\|x\| \geq 2M$  we have  $\|y\| = \|x - t\|x\|\xi\| \geq (1-t)\|x\| \geq M$ . Then for  $y \in W(x)$  and  $\|x\| \geq 2M$ ,

$$\begin{aligned} \langle \xi, m(y) \rangle &= \langle (\xi - n(x)) + (n(x) - n(y)) + n(y), m(y) \rangle \\ &\leq \|\xi - n(x)\| + \|n(x) - n(y)\| + \langle n(y), m(y) \rangle < 0. \end{aligned}$$

Hence (3.10) holds, since if  $y \in W(x)$  and  $\|x\| \geq 2M$ , then

$$\log \frac{\pi(y)}{\pi(x)} = \log \pi(y) - \log \pi(x) = -\|x\| \int_0^t \langle \xi, \nabla \log \pi(x - s\|x\|\xi) \rangle ds > 0.$$

Next we prove

$$\liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{\pi(x^*)}{\pi(x)} > 0 \right) \leq \liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{q(x^*)}{q(x)} > 0 \right). \quad (3.11)$$

By the RWM property of the MpCN kernel (Proposition 2.1), the law of  $\xi(x)/2 = \log(\|x^*\|/\|x\|)$  does not depend on  $x$ . Therefore for any  $\delta > 0$  there exists  $C > 0$  such that

$$\mathbb{P}_x \left( d \log \frac{\|x^*\|}{\|x\|} \leq -C \right) < \delta.$$

Therefore by (3.7)

$$\begin{aligned} \liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{\pi(x^*)}{\pi(x)} > 0 \right) &= \liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{q(x^*)}{q(x)} > d \log \frac{\|x^*\|}{\|x\|} \right) \\ &\leq \liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{q(x^*)}{q(x)} > -C \right) + \mathbb{P}_x \left( d \log \frac{\|x^*\|}{\|x\|} \leq -C \right) \\ &\leq \liminf_{x \rightarrow \infty} \mathbb{P}_x \left( \log \frac{q(x^*)}{q(x)} > 0 \right) + \delta. \end{aligned}$$

Hence (3.11) is proved. Thus for  $\|x\| \geq 2M$ , by (3.10) and (3.11),

$$\begin{aligned} P(x, \{x\}^c) &= \mathbb{E}_x \left[ \min \left\{ 1, \frac{q(x^*)}{q(x)} \right\} \right] \\ &\geq \mathbb{P}_x \left( \log \frac{q(x^*)}{q(x)} > 0 \right) \\ &\geq \mathbb{P}_x \left( \log \frac{\pi(x^*)}{\pi(x)} > 0 \right) + o(\|x\|) \\ &\geq \mathbb{P}_x (x^* \in W(x)) + o(\|x\|). \end{aligned}$$

Observe that  $UW(x) = \{Uy; y \in W(x)\} = W(Ux)$  for any unitary matrix  $U$ , and  $x \in W(x) \Leftrightarrow e \in W(e)$  for  $e = x/\|x\|$ . By these facts

$$\mathbb{P}_x (x^* \in W(x)) = \mathbb{P}_e (x^* \in W(e)) > 0$$

for any  $e \in S^{d-1}$ . Thus  $\liminf_{x \rightarrow \infty} P(x, \{x\}^c) = \Pi\text{-ess sup } P(x, \{x\}) - 1 \geq \mathbb{P}_e (x^* \in W(e)) > 0$ . Thus the claim follows by Proposition 3.5.  $\square$

## Acknowledgement

The author would like to extend thanks to Alexandros Beskos for interesting discussions related to the similarity of MpCN and MMALA algorithms. I also thank to Masayuki Uchida for fruitful discussions related to practical implementation of MpCN for inference of discretely observed stochastic diffusion process.

## References

- Yves Atchadé and Gersende Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 16(1):116–154, 2010. ISSN 1350-7265. doi: 10.3150/09-BEJ199. URL <http://dx.doi.org/10.3150/09-BEJ199>.
- Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8(3):319–350, 2008. ISSN 0219-4937. doi: 10.1142/S0219493708002378. URL <http://dx.doi.org/10.1142/S0219493708002378>.
- N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989. ISBN 0-521-37943-1.
- Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Handbooks of modern statistical methods. 2011.
- Petros Dellaportas and Gareth O. Roberts. An introduction to MCMC. In *Spatial statistics and computational methods (Aalborg, 2001)*, volume 173 of *Lecture Notes in Statist.*, pages 1–41. Springer, New York, 2003.
- Randal Douc, Gersende Fort, Eric Moulines, and Philippe Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004. ISSN 1050-5164. doi: 10.1214/105051604000000323. URL <http://dx.doi.org/10.1214/105051604000000323>.
- S. Dutta. Multiplicative random walk Metropolis-Hastings on the real line. *ArXiv e-prints*, August 2010.
- Gersende Fort and Eric Moulines.  $V$ -subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Probab. Lett.*, 49(4):401–410, 2000. ISSN 0167-7152. doi: 10.1016/S0167-7152(00)00074-2. URL [http://dx.doi.org/10.1016/S0167-7152\(00\)00074-2](http://dx.doi.org/10.1016/S0167-7152(00)00074-2).
- Gersende Fort and Eric Moulines. Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.*, 103(1):57–99, 2003. ISSN 0304-4149. doi: 10.1016/S0304-4149(02)00182-5.
- Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014. ISSN 1050-5164. doi: 10.1214/13-AAP982. URL <http://dx.doi.org/10.1214/13-AAP982>.
- Søren F. Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85(2):341–361, 2000. ISSN 0304-4149. doi: 10.1016/S0304-4149(99)00082-4.
- Søren F. Jarner and Gareth O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.*, 12(1):224–247, 2002. ISSN 1050-5164. doi: 10.1214/aoap/1015961162.
- Søren F. Jarner and Gareth O. Roberts. Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.*, 34(4):781–815, 2007. ISSN 0303-6898. doi: 10.1111/j.1467-9469.2007.00557.x. URL <http://dx.doi.org/10.1111/j.1467-9469.2007.00557.x>.
- Søren F. Jarner and Richard L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli*, 9(4):559–578, 2003. ISSN 1350-7265. doi: 10.3150/bj/1066223269.
- Leif T. Johnson and Charles J. Geyer. Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.*, 40(6):3050–3076, 2012. ISSN 0090-5364. doi: 10.1214/12-AOS1048.
- Kengo Kamatani. Metropolis-Hastings algorithms with acceptance ratios of nearly 1. *Ann. Inst. Statist. Math.*, 61(4):949–967, 2009. ISSN 0020-3157. doi: 10.1007/s10463-008-0180-6.

- Kengo Kamatani. Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Arxiv*, 2014. URL <http://arxiv.org/abs/1412.6231>.
- S. Livingstone. Geometric ergodicity of the Random Walk Metropolis with position-dependent proposal covariance. *ArXiv e-prints*, July 2015.
- Kerrie L. Mengersen and Richard L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996. ISSN 0090-5364. doi: 10.1214/aos/1033066201.
- Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73182-9. doi: 10.1017/CBO9780511626630. With a prologue by Peter W. Glynn.
- Esa Nummelin. *General irreducible Markov chains and nonnegative operators*. Number 83 in Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1984.
- Sidney I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.
- Gareth O. Roberts and Richard L. Tweedie. Geometric convergence and central limit theorems for multi-dimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996. ISSN 0006-3444. doi: 10.1093/biomet/83.1.95.
- Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. ISSN 0090-5364. doi: 10.1214/aos/1176325750. With discussion and a rejoinder by the author.
- Pekka Tuominen and Richard L. Tweedie. Subgeometric rates of convergence of  $f$ -ergodic Markov chains. *Adv. in Appl. Probab.*, 26(3):775–798, 1994. ISSN 0001-8678. doi: 10.2307/1427820.